基于双通道特征融合的 WPOS-GRU 专利分类方法 *

余本功[a, b]，张培行[a]

(合肥工业大学 a. 管理学院; b. 过程优化与智能决策教育部重点实验室, 合肥 230009)

摘 要：为提高专利文本的自动分类的效率和准确度，提出一种双通道特征融合 WPOS-GRU(word2vec and part of speech gated recurrent unit)专利文本自动分类方法。首先获取专利摘要文本，并进行清洗和预处理；然后对专利文本进行词向量表示和词性标注，并将专利文本分别映射为 word2vec 词向量序列和 POS 词性序列；最后使用两种特征通道训练 WPOS-GRU 模型，并对模型效果进行实验分析。通过对比传统专利分类方法和单通道专利分类方法，双通道特征融合的 WPOS-GRU 专利分类方法提高了分类效果。本文提出的方法节省了大量的人力成本，提高了专利文本分类的准确度，更能满足大量专利文本分类任务自动化高效率的需要。

关键词：专利分类；词性标注；特征融合；门限递归单元

中图分类号：TP 391 doi: 10.3969/j.issn.1001-3695.2018.08.0628

WPOS-GRU patent classification method based on dual channel feature fusion

Yu Bengong[a,b], Zhang Peihang[a]

(1. *School of Management*, b. *Key Laboratory of Process Optimization & Intelligent Decision-Making* of *Ministry of Education*, *Hefei University of Technology*, *Hefei* 230009, *China*)

Abstract: In order to improve the efficiency and accuracy of patent text automatic classification, this paper proposes a two-channel feature fusion WPOS-GRU patent text automatic classification method. Firstly, the patent summary text is obtained, cleaned and pretreated, then the patent text is represented by word vector and part-of-speech tagging, and the patent text is mapped into word 2vec word vector sequence and POS part-of-speech sequence, respectively. Finally, WPOS-GRU model is trained by two feature channels, and the effect of the model is analyzed experimentally. By comparing the traditional patent classification method with the single-channel patent classification method, the WPOS-GRU patent classification method with two-channel feature fusion improves the classification effect. The method proposed in this paper saves a lot of manpower costs, improves the accuracy of patent text classification, and can meet the needs of automation and high efficiency of a large number of patent text classification tasks.

Key words: patent classification; part of speech tagging; feature Fusion; GRU

0 引言

近年来，科技创新越来越引起人们的重视，而专利作为创新的重要记录载体，也呈现出爆炸增长的态势[1]。据在专利申请领域，在 2016 年的一年时间内，中国共受理专利申请 1 30 万件，且呈逐年上升趋势。如此海量的专利数据，通过人工分类需要巨大的人力成本，处理效率无法满足实际需要，专利的自动分类方法研究的重要性日渐突显，已成为现阶段一个重要的研究热点问题。目前，专利分类研究多采用 IPC 分类体系，IPC 分类是一种层次结构分类体系，包括部、大类、小类、组等层次，是世界上使用较多普遍认可的一种分类体系。目前，相关研究者通过机器学习来处理专利自动分类问题，通过进行文本分析，提取文本中关键的特征词，并结合机器学习分类器完成分类，取得不错的效果。最近几年，深度学习在自然语言处理领域取得了很好的效果，且端到端的处理流程更能满足专利自动分类的需要，使用深度学习模型实现专利的自动分类是一种较好的解决思路。

1 相关工作

专利文本是科技创新的一种重要表现形式，对专利的分析和利用一直被世界各国所重视，专利的自动分类，对专利审核、专利检索有着非常重要的意义。使用机器学习进行专利分类一直是该问题的研究热点。李生珍等人[2]对文本进行分词并提取特征词，将专利文本映射成特征向量，并使用 BP 神经网络构建分类器，马芳等人[3]使用径向基函数神经网络构建分类模型，并设计了专利自动分类系统。相比于普通文本，专利文本具有其自身的殊性，有针对的构建分类器，更能适应专利自动分类的需要[4]。屈鹏等人[5]认为专利文本有较明显的专业特征，使用专业术语构建特征能提高专利分类的效果，并使用朴素贝叶斯、SVM 等分类器进行实验，对比了机器学习相关分类器的效果。基于向量空间模型的分类方法，忽略了词语间的语义信息，廖列法等人[6]认为用主题代替传统的向量空间模型，在构建分类器时考虑到了语义信息。

深度学习近些年的飞速发展，为自然语言处理很多问题提供了新的解决思路，尤其在文本分类问题上表现出良好的性能。一些学者通过自动编码机来处理特征，提取出文本中深层次的信息[7]，受此启发，马双刚[8]将自动编码机应用在专利自动分类中，并取得了不错的效果。目前相关研究者对专利自动分类的研究多集中在特征提取和处理上，而端到端和深度神经网络可以摆脱特征工程的束缚，更适合大量专利

收稿日期：2018-08-25；**修回日期：**2018-10-23 **基金项目：**国家自然科学基金资助项目（71671057）

作者简介：余本功（1971-），男，安徽合肥人，教授，博士，主要研究方向为信息系统、知识管理（bgyu19@163.com）；张培行（1993-），男，山东滕州人，硕士，主要研究方向为信息管理与信息系统、文本挖掘、机器学习.

数据自动分类的需要[9]。在深度学习模型中，首先要进行 word2vec 词向量训练，将词映射成一个低维的向量，解决了传统词袋模型词向量维度过大的难题[10]，且 word2vec 向量的训练过程结合了词的上下文内容，包含了词的语义信息，在深度学习研究中应用较多。Kim 等人[11]使用卷积神经网络构建文本分类器，提取文本深层次的特征，不需要人工干预，相比于传统人工提取特征的方法，更高效快捷，在分类效果上也更优于传统方法。胡杰等人[12]使用卷积神经网络进行特征提取，并使用随机森林构造分类器对专利文本进行分类，通过对比随机森林等方法，该方法较好的提升了分类效果。一些研究者认为，卷积神经网络有着局部联接的特点，所以在提取特征的过程中，会忽略文本的结构特征，而 LSTM 和 GRU 等循环神经网络模型是一种序列模型，更适合文本特征的提取[13, 14]。王树恒等人[15]使用双向的 LSTM 模型对文本情感进行分类，通过实验 LSTM 获得了比 CNN 更好的分类准确率。李雪莲等人[16]通过对比实验分析了 LSTM 和 GRU 模型结构和性能，并指出 GRU 模型 继承了 LSTM 自动学习的功能，但其结构更为简单，大大缩短了模型训练时间，更适合大量文本数据的研究应用。深度学习方法应用在舆情发现、情感分析等方面取得了很好的成绩，而鲜有研究者将深度学习方法用于专利自动分类领域。本文分析了专利文本的特点，提取专利摘要文本，结合 LSTM 深度学习模型完成专利摘要文本的自动分类。此外，每个词都有不同的词性，词性包含了一些重重要的语义内容[17]，尤其在专利文本中，如名词的重要性一般较高，而 word2vec 词向量模型忽略了词性信息，所以本文在 GRU 的基础上，将文本的 word2vec 词向量和词性信息进行结合，实现语义和词性双通道建模，提出了双通道特征融合的 WPOS-GRU 专利摘要自动分类方法。

2　相关关键技术

2.1　word2vec 词向量训练和 POS 词性标注

word2vec 模型由 Mikolov 等人研发的词向量表示工具，包括 CBOW 模型和 Skip-gram 模型，本文使用 CBOW 模型实现词向量训练。CBOW 模型结构如图 1 所示，包括输入层、映射层和输出层。

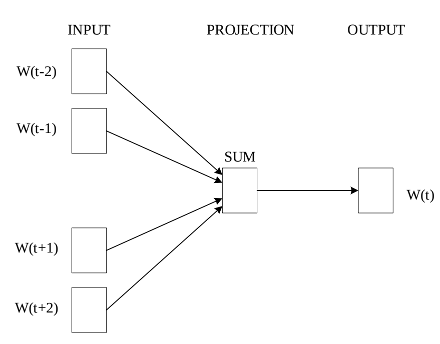


图 1　Skipgram 模型结构图

Fig. 1　Skipgram model structure diagram

输入层为一个滑动窗口，将该词上下文的 n 个词向量输入到模型，输出当前词的向量表示。因为其在词向量表示时，考虑到词的上下文信息，所以最终的词向量表示了一定的语义信息，可以通过向量距离计算来求两个词的相似度。

词性标注是自然处理中很重要的工作，主要包括基于规则的方法、基于统计的方法和机器学习方法，常被用于机器翻译、文字识别等领域，为每个词赋予其对应的词性，作为后面语义分析的基础工作。本文使用的词性标注标准为中科院计算所标注集，包括名词、时间词、处所词、方位词、动词、形容词等类别。该标注标准标注信息较为详细，在相关研究中应用较广。

2.2　GRU 循环神经网络模型

GRU 循环神经网络模型在传经循环神经网络中加入门的因素，可以保留更长距离的记忆，相比于 LSTM，其减少了门的数量，并将 LSTM 隐层状态和细胞状态合并，减少了冗余信息，所以其模型效果和 LSTM 相似，而由于参数的减少，计算效率大大提高。GRU 包括更新门和控制门，是控制信息选择性通过的机制，由一个 sigmoid 神经网络层和一个向量点乘组成。门元素的值在[0,1]，当值为 1 时表示信息完全通过，当值为 0 时表示信息完全阻塞。GRU 结构信息如图 2 和式(1)所示。

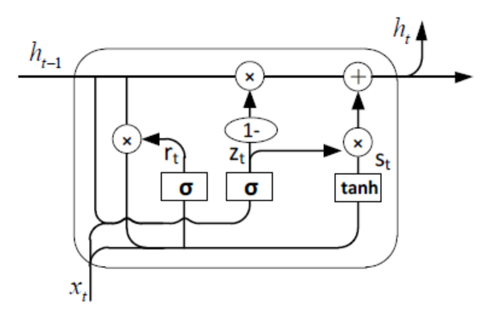


图 2　GRU 节点结构图

Fig. 2　GRU node structure diagram

$$
\begin{aligned}
z_t &= \sigma(U_z x_t + W_z h_{t-1} + b_z) \\
r_t &= \sigma(U_r x_t + W_z h_{t-1} + b_z) \\
s_t &= \tanh(U_s x_t + W_s \cdot r_t * h_{t-1} + b_s) \\
h_t &= (1 - z_t) * h_{t-1} + z_t * s_t
\end{aligned} \tag{1}
$$

其中：z_t 代表更新门，用来控制当前输入所占的比重，R_t 为重置门，用来控制上一步的记忆中哪些对当前输入起作用。W_z, W_r, W_s 代表权重,b_z b_r b_s 代表偏置量。x_t 代表 t 时刻的输入，s_t 代表需要更新的信息，h_t 代表 t 时间步的隐藏层状态，σ 代表非线性函数。

3　基于双通道特征融合的 WPOS-GRU 专利摘要的自动分类模型

本文首先对文本数据进行预处理，包括数据清洗、分词、去停用词等工作，然后对文本进行词向量表示和词性标注，并将词向量和标注信息输入进双通道 GRU 循环神经网络模型，完成训练，最后使用该模型对文本进行分类测试。

3.1　专利数据获取及数据预处理

专利数据包括题目、摘要、正文、主分类号等文本项，其中摘要文本中包含了该篇专利的核心内容，阅读者通过阅读专利摘要就可以对该专利的类别有所把握，而阅读全文则需要耗费大量时间成本，所以本文需要的获取专利的摘要文本信息和主分类号。

对专利摘要文本进行数据清洗，除去因网络来源产生的噪声，然后进行分词和词性标注处理，词性标注采用中科院计算所的标注标准。对分词和标注后的结果进行向量表示，将词和标注信息分别映射成低维向量。

3.2　专利文本分类模型

在专利摘要文本中，相比于其他文本数据，语言较为领域化专业化，且其专业术语较广，传统基于特征词的方法无法对术语词进行很好的覆盖，一旦有新的术语词出现，特征向量就要重新设计。所以本文提出基于循环神经网络的方法，其方法通用性较强，新术语词往往是由旧词组成的短语，通过旧词的词向量计算，可以得出新短语的 word2vec 语义。无

须人工提取特征，适用专利摘要文本特征词提取不方便的特征，节约了大量的人力成本，更适应大量专利数据自动分类的需要。本文考虑到专利文本中，不同的词性所代表的重要性不同，而 word2vec 词向量忽略了词性信息，所以本文在 GRU 的基础上，提出了融合词性信息的专利分类模型 WPOS-GRU。其中词性信息包含了一定的语义信息，可以单独输入模型，也可以作为 word2vec 向量的补充内容，所以本文设计了 word2vec 单通道特征 GRU、POS 单通道特征 GRU 和双通道融合特征 WPOS-GRU 三种模型结构,并进行对比实验。

3.2.1 双通道融合特征 WPOS-GRU 模型

双通道特征融合 WPOS-GRU 模型包括 word2vec 和 POS 两个通道，其模型结构如图 3 所示。

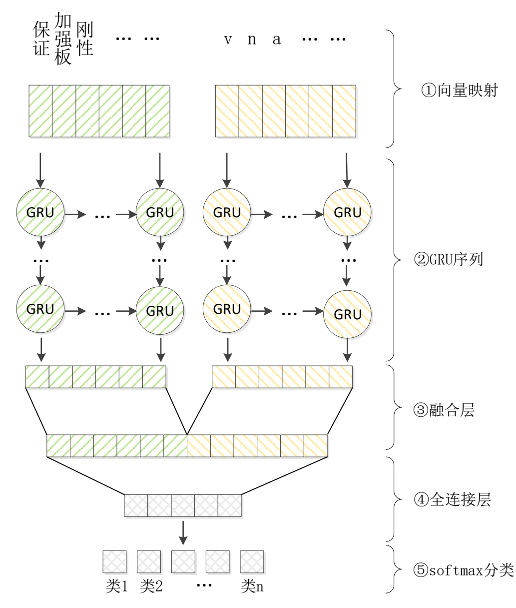


图 3　双通道特征融合 WPOS-GRU 结构

Fig. 3　Dual channel feature fusion WPOS-GRU structure

①表示映射层，即将摘要中的词映射成向量的形式。仍以“保证加强板刚性结构，使其具有良好的承载能力和抗机械冲击能力”为例，将“保证、加强”等词分别映射成 word2vec 词向量和 POS 词性向量。

②表示 GRU 序列层。将上一层映射后的 word2vec 词向量序列和 POS 词性向量序列分别按照词序输入到第一层 GRU 的每一个时间步中，每层 GRU 的输出结果作为下一层 GRU 的输入，最后一层 GRU 节点只保留最后一个节点的输出。其中 word2vec 词向量特征通道的 GRU 序列输出为 O1,POS 词性向量特征通道的 GRU 序列输出为 O2。

③表示融合层，将 word2vec 词向量通道深层 GRU 输出向量 O1 和 POS 词性特征通道 GRU 输出向量 O2 进行拼接，拼接后的向量表示为 O。

④表示全连接层，假设该专利数据集共包含 n 个类别，全连接层的节点数为 n，将融合层的输出向量映射成 n 维向量 。

⑤表示 softmax 分类层，对上一层的输出向量进行规一化，得到新的 n 维向量，其中每个元素代表属于该类别的概率。

其模型训练过程如下：

多通道特征融合 WPOS-GRU 算法流程

输入：训练集 $D=\{(x_k,p_k,y_k)\}^m_{k=1}$，超参数：θ。

初始化训练参数 W

repeat

for all $(x_k,p_k,y_k)\in D$ do

1 计算当前样本的输出 y'_k

2 计算输出 y'_k 和样本标签 y 的差值 E

3 根据 E 计算各参数的梯度

4 更新各参数的值

end for

until:训练误差收敛到一个比较小的值。

输出：各训练参数确定的分类网络。

其中：D 为 m 篇专利摘要训练集，x_k 表示第 k 篇专利摘要的词向量表示序列，p_k 表示第 k 篇文档的词性向量表示序列，y_k 为第 k 篇文档的类别向量，y'_k 表示第 k 篇文本预测类别向量。

3.2.2 单通道特征 GRU 模型

单通道特征 GRU 模型只有一个通道，word2vec 通道或者 POS 通道结构如图 4 所示。

①表示映射层，即将摘要中的词映射成向量的形式。以“保证加强板刚性结构，使其具有良好的承载能力和抗机械冲击能力”为例，将“保证、加强”等词映射成向量形式，在 word2vec 单通道特征 GRU 模型中，将词映射成 word2vec 向量；在 POS 单通道特征 GRU 模型中，将词映射成 POS 词性标注向量。

②表示 GRU 序列层。将上一层映射后的向量序列按照词序步输入到第一层 GRU 的每一个时间步中，每层 GRU 的输出结果作为下一层 GRU 的输入，最后一层 GRU 节点只保留最后一个节点的输出。

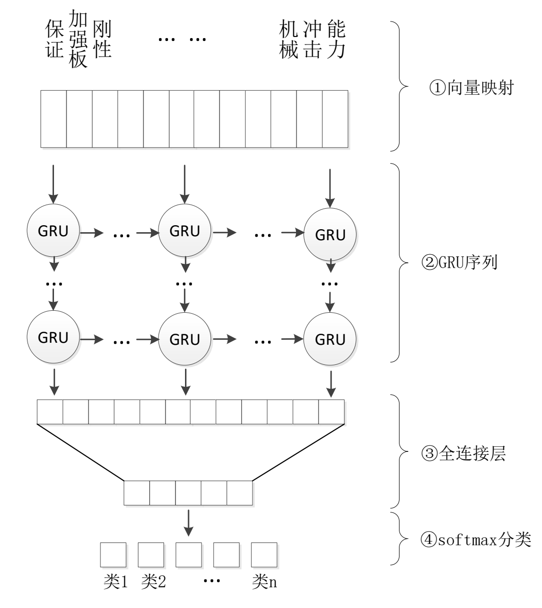


图 4　单通道特征 GRU 模型结构

Fig. 4　Single channel characteristic GRU model structure

③表示全连接层，假设该专利数据集共包含 n 个类别，全连接层的节点数为 n，将 GRU 序列层的输出映射成 n 维向量 。

④表示 softmaxt 分类层，对上一层的输出向量进行规一化，得到新的 n 维向量，其中每个元素代表属于该类别的概率。

其模型训练过程如下：

word2vec 单通道特征 GRU 算法流程

输入：训练集 $D=\{(x_k, y_k)\}^m_{k=1}$，超参数：θ。

初始化训练参数 W

repeat

for all $(x_k, y_k)\in D$ do

1 计算当前样本的输出 y'_k

2 计算输出 y'_k 和样本标签 y 的差值 E

3 根据 E 计算各参数的梯度

4 更新各参数的值

end for

until:训练误差收敛到一个比较小的值。

输出：各训练参数确定的分类网络。

其中：D 为 m 篇专利摘要训练集，x_k 表示第 k 篇专利摘要的词向量表示序列，y_k 为第 k 篇文档的类别向量，y'_k 表示第 k 篇文本预测类别向量。

POS 单通道特征 GRU 算法流程

输入：训练集 $D=\{(p_k, y_k)\}^m_{k=1}$，超参数：θ。

初始化训练参数 W

repeat

for all $(p_k, y_k)\in D$ do

1 计算当前样本的输出 y'_k

2 计算输出 y'_k 和样本标签 y 的差值 E

3 根据 E 计算各参数的梯度

4 更新各参数的值

end for

until:训练误差收敛到一个比较小的值。

输出：各训练参数确定的分类网络。

其中：D 为 m 篇专利摘要训练集，x_k 表示第 k 篇专利摘要的词性向量序列，y_k 为第 k 篇文档的类别向量，y'_k 表示第 k 篇文本预测类别向量。

3.3　模型评价

为设计对比实验，评价本文方法的可行性，本文首先对所有摘要数据划分训练集和验证集。使用上述算法完成模型训练，然后使用精度、准确率、召回率和 F1 值来对模型效果进行评价。

精度 accuracy 指的是正确分类的专利文本数 T 和专利文本总数 N 的比值，如式（2）所示。

$$accuracy=\frac{T}{N} \tag{2}$$

准确率 Precision 指的是预测为该类别的专利文本中，实际属于该类别的比例，如式（3）所示，其中 TP 表示预测为该类别且预测正确的文档数，FP 表示预测为该类别且预测错误的文档数。

$$precision=\frac{TP}{TP+FP} \tag{3}$$

召回率 Recall 指的是实际属于该类别的专利文档中，被预测出来的比例，如式（4）所示，其中 TP 表示预测为该类别且预测正确的文档数，FN 表示实际属于该类别但预测错误的文档数。

$$recall=\frac{TP}{TP+FN} \tag{4}$$

F1 值同时兼顾了准确率和召回率，如式（5）所示。

$$F1=\frac{2\cdot precision\cdot recall}{precision+recall} \tag{5}$$

4　实验分析

4.1　专利数据获取及数据预处理

本文在进行实验时，选择了作者较为熟悉的计算机领域，参考文献[8]选择专利的方法，进行专利文本的分类实验。首先在 SooPAR 专利搜索引擎上选择中国专利“计算机”主题，专利类型限定为发明专利，检索出和计算机相关的专利文本的主分类号有 G06F1/16、G06F1/18、G06F1/20、G06F3/02、G06F3/14 五个类别，类别粒度为小组级别。然后在上海知识产权上平台上检索这五个类别的专利，为保证时效性，选择了近 4 年的专利数据，每个类别选择 2000 篇进行下载。对下载后的专利数据进行筛选去重，保留专利摘要文本并分类存储。

对摘要文本进行分词和去停用词处理。本实验使用 jieba 分词工具，该工具分词效果较好，相关研究中使用较多。

因 word2vec 训练需要大文本语料集，所以本文将维基百科文本和专利文本进行组合，其中维基百科库为网络公开数据，大小约 1.3 GB，专利数据共 10 000 条。通过组合，既满足大文本语料的要求，也包含了领域信息，保证了词向量训练的效果。最后使用 word2vec 模型将评论文本中的词映射成 100 维向量。

本文词性标注使用 jieba 工具包，标注标准为中科院计算所的标注集，该标注集较多全面，能为自然语言处理相关工作提供辅助，在相关研究中使用较多。对词性标注内容进行 onehot 编码，编码成长度为 50 的向量，如形容词 a 标注为[0,0,1,0,0,…]，其中 a 对应向量中第三个位置。

4.2　模型训练

本文实验环境配置如表 1 所示。

表 1　实验环境参数

Table 1　Experimental environmental parameters

参数	值
处理器	Intel(R) Core™　i5-7300 HQ CPU @ 2.50GHz
内存	8GB
显卡	NVIDIA GeForce GTX 1050
编程语言	Python 3
深度学习库	Tensorrflow + Keras

GRU 循环神经网络序列长度为句子长度，GRU 节点数取 50，100，150 时取得较高的 F1 值，由于专利文本分类需要较高效率，所以本文节点数定为 50。

本文将五个类别的 10000 专利文本划分训练集和验证集，其中训练集为 8 000 条，验证集为 2 000 条。模型训练过程中，使用交叉熵作为损失函数，使用 Adam 方法作为优化函数，模型训练三个 epoch 后收敛。

为验证本文方法的有效性，分别训练本文三个模型，并设计对比实验，将本文方法与专利分类中常用的 NB（朴素贝叶斯）、SVM（支持向量机）、NN（神经网络）和 RF（随机森林）相对比，实验结果如表 2 和图 5 所示。此外，为验证不同词性的重要程度不同，本文使用仅名词、仅形容词、仅动词的词向量训练 word2vec 单通道特征 GRU，得出评价结果。

通过对比实验可以看出，在传统机器学习模型中,神经网络表现最好，分类精度为 0.92；本文提取的三种方法中，word2vec 单通道特征 GRU 分类精度为 0.95，相比传统方法提升较大，所以将深度学习方法应用到专利分类中是有重要价值的；训练单通道特征 GRU，若仅使用名词训练，分类精度为 0.91，使用动词训练，分类精度为 0.81，而仅使用形容词训练，分类精度仅为 0.53，说明不同词性的词对分类的贡献是不同的，名词包含的信息量最大；此外，pos 单通道特征 GRU 分类精度为 0.46，在五分类实验中，随机值为 0.2，说明使用词性特征包含了较多的语义信息，双通道融合后模型分类精度在 0.974，能够完成专利文本自动分类任务。从

F1 上看，双通道特征融合 WPOS-GRU 效果也较好，相比传统方法和单通道特征也有较大提升。

5　结束语

专利作为科技创新的一种重要表现形式，专利文本的分类，对于专利的分析利用有很重要的作用。本文结合深度学习技术，就大量专利文本的自动高效的分类问题，提出了双通道特征融合 WPOS-GRU 专利自动分类模型，该模型通过引入词性语义信息，提高了专利文本自动分类的准确度，使得专利自动分类结果更可靠实用。但本文方法仍存在一些不足，如一些新生类别下缺少专利文本数据，本文方法表现可能不佳。

表 2　对比实验结果

Table 2　Contrast experiment results

	acc	pre	recall	f1
NB	0.8177	0.8238	0.8087	0.8162
SVM	0.8374	0.8453	0.7639	0.8025
RF	0.9078	0.9631	0.8799	0.9196
NN	0.9225	0.9533	0.9163	0.9345
pos	0.4664	0.4857	0.6292	0.5482
word2vec	0.9501	0.9322	0.9423	0.9372
名词	0.9161	0.9141	0.9101	0.9121
动词	0.8365	0.8357	0.8044	0.8198
形容词	0.5778	0.5907	0.4927	0.5373
WPOS-GRU	0.9740	0.9707	0.9671	0.9689

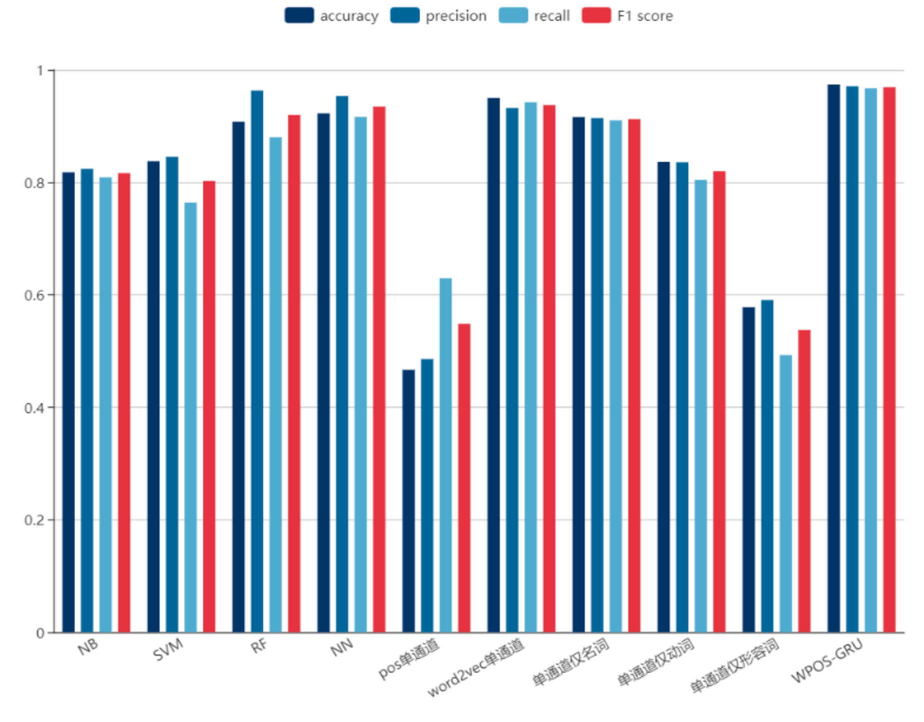


图 5　对比实验结果

Fig. 5　Contrast experiment results4.3 实验结果分析

参考文献：

[1] WIPO. World intellectual property report 2017 [EB/OL].（2017-12-06）. http://www. wipo. int/edocs/pubdocs/en/wipo_pub_944_2017. Pdf.

[2] 李生珍，王建新，齐建东，等. 基于 BP 神经网络的专利自动分类方法 [J]. 计算机工程与设计, 2010, 31(23): 5075-5078. (Li Shengzhen, Wang Jianxin, Qi Jiandong, *et al.* Automated categorization of patent based on back-propagation network [J]. Computer Engineering and Design. 2010, 31(23): 5075-5078.)

[3] 马芳. 基于 RBFNN 的专利自动分类研究 [J]. 现代图书情报技术, 2011(12): 58-63. (Ma Fang. Research of patent automatic classification based on RBFNN [J]. New Technology of Library and Information Service, 2011(12): 58-63.)

[4] 刘红光，马双刚，刘桂锋. 基于机器学习的专利文本分类算法研究综述 [J]. 图书情报研究，2016(3): 79-86. (Liu Hongguang, Ma Shuanggang, Liu Guifeng. A review of research on patent document classification algorithms based on machine learning [J]. Library and Information Studies, 2016 (3): 79-86.)

[5] 屈鹏，王惠临. 专利文本分类的基础问题研究 [J]. 现代图书情报技术, 2013, 29(3): 38-44. (Qu Peng, Wang Huilin. Fundamental research questions in patent text categorization [J]. New Technology of Library and Information Service, 2013, 29(3): 38-44.)

[6] 廖列法，勒孚刚，朱亚兰. LDA 模型在专利文本分类中的应用 [J]. 现代情报, 2017, 37(3): 35-39. (Liao Liefa, Le Fugang, Zhu Yalan. The application of LDA model in patent text classification [J]. Journal of Modern Information, 2017, 37(3): 35-39.)

[7] 刘勘，袁蕴英. 基于自动编码器的短文本特征提取及聚类研究[J]. 北京大学学报:自然科学版，2015,51(2): 282-288. (Liu Kan, Yuan Yunying. Short Texts Feature Extraction and Clustering Based on Auto-Encoder [J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2015,51(2): 282-288.)

[8] 马双刚. 基于深度学习理论与方法的中文专利文本自动分类研究 [D]. 镇江：江苏大学，2016.(Ma Shuanggang.Study of automatic Chinese patent classification based on deep learning theory and method[D].Zhenjiang: Jiangsu University,2016.)

[9] Tajbakhsh N,Suzuki K. Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs [J]. Pattern Recognition, 2017, 63(3): 476-486

[10] Mikolov T, Chen K, Corrado G, *et al.* Efficient estimation of word representations in vector space [J]. Computer Science, 2013.

[11] Kim Y. Convolutional neural networks for sentence classification [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg,PA: Association for Computational Linguistics, 2014: 1746-1751.

[12] 胡杰，李少波，于丽娅，等. 基于卷积神经网络与随机森林算法的专利文本分类模型[J]. 科学技术与工程，2018,18(6) . (Hu Jie, Li Shaobo, Yu Liya, *et al.* A patent classification model based on convolutional neural networks and rand forest [J]. Science Technology and Engineering, 2018, 18(6) , 268-272.)

[13] Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures[J]. Journal of International Neural Network Society, 2005, 18(5-6): 602.

[14] Chung J, Gulcehre C, Cho K, *et al*. Gated Feedback recurrent neural networks [J]. Computer Science, 2015: 2067-2075.

[15] 王树恒，吐尔根・依布拉音，卡哈尔江・阿比的热西提,等.基于 BLSTM 的维吾尔语文本情感分析 [J]. 计算机工程与设计, 2017, 38 (10):2879-2886. (Wang Shuheng, Turgun Lbrahim, Kahaerjiang Abiderexiti, *et al.* Sentiment classification of Uyghur text based on BLSTM [J]. Computer Engineering and Design, 2017, 38(10): 2879-2886.)

[16] 李雪莲，段鸿，许牧. 基于门循环单元神经网络的中文分词法 [J]. 厦门大学学报:自然科学版,2017，56(2): 237-243. (Li Xuelian, Duan Hong, Xu Mu. A Gated Recurrent Unit Neural Network for Chinese Word Segmentation [J]. Journal of Xiamen University, 2017，56(2): 237-243.)

[17] Tang Xiaoyan, Cao Jing. Automatic genre classification via n-grams of part-of-speech tags [J]. Procedia-Social and Behavioral Sciences, 2015, 198(7): 474-478..